

2/5/1 (Item 1 from file: 347)
DIALOG(R)File 347:JAPIO
(c) 2000 JPO & JAPIO. All rts. reserv.

04764308 **Image available**
DOCUMENT PROCESSOR

PUB. NO.: 07-056908 JP 7056908 A]
PUBLISHED: March 03, 1995 (19950303)
INVENTOR(s): MURATA MAKOTO
APPLICANT(s): FUJI XEROX CO LTD [359761] (A Japanese Company or
Corporation), JP (Japan)
APPL. NO.: 05-218142 [JP 93218142]
FILED: August 09, 1993 (19930809)
INTL CLASS: [6] G06F-017/21; G06F-017/27; G06F-017/30
JAPIO CLASS: 45.4 (INFORMATION PROCESSING -- Computer Applications)

ABSTRACT

PURPOSE: To provide a document processor which can efficiently match structure patterns.

CONSTITUTION: A logical structure holding part 1 holds the logical structure of a document. A template holding part 2 holds the template of the document. The logical structure of the document held in the logical structure holding part 1 has structure based on the template held in the template holding part 2. A structure pattern holding part 3 holds the structure pattern which is matched. A structure pattern matching processing part 4 restricts a range where a retrieval processing is executed among the logical structures of the document, which are held in the logical structure holding part 1, based on the template of the document, which is held in the template holding part 2, and the structure pattern held in the structure pattern holding part 3. A part fitted to the structure pattern is retrieved within the restricted range and among the logical structures of the document.

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-56908

(43) 公開日 平成7年(1995)3月3日

(51) Int.Cl.⁸

識別記号

庁内整理番号

F I

技術表示箇所

G 0 6 F 17/21
17/27
17/30

7315-5L
7315-5L

G 0 6 F 15/ 20

5 9 0 E
5 5 0 E

審査請求 未請求 請求項の数 1 F D (全 5 頁) 最終頁に続く

(21) 出願番号

特願平5-218142

(22) 出願日

平成5年(1993)8月9日

(71) 出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂三丁目3番5号

(72) 発明者 村田 真

神奈川県横浜市保土ヶ谷区神戸町134番地

横浜ビジネスパークイーストタワー 富

士ゼロックス株式会社内

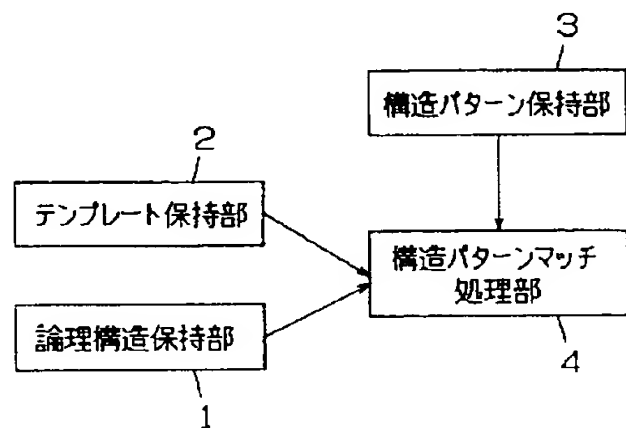
(74) 代理人 弁理士 石井 康夫 (外1名)

(54) 【発明の名称】 文書処理装置

(57) 【要約】

【目的】 効率よく構造パターンマッチを行なうことのできる文書処理装置を提供する。

【構成】 論理構造保持部1は、文書の論理構造を保持する。テンプレート保持部2は、文書のテンプレートを保持する。論理構造保持部1に保持されている文書の論理構造は、テンプレート保持部2に保持されているテンプレートに基づいた構造を有している。構造パターン保持部3は、構造パターンマッチを行なう構造パターンを保持している。構造パターンマッチ処理部4は、テンプレート保持部2に保持されている文書のテンプレートと、構造パターン保持部3に保持されている構造パターンに基づいて、論理構造保持部1に保持されている文書の論理構造のうちから、探索処理を行なう範囲を限定する。そして、限定された範囲内で、文書の論理構造のうちから、構造パターンに適合する部分を探索する。



【特許請求の範囲】

【請求項1】 構造化文書の検索、自動生成を行なう文書処理装置において、文書の論理構造のテンプレートを保持するテンプレート保持部と、文書の論理構造を保持する論理構造保持部と、構造パターンを保持する構造パターン保持部と、前記論理構造保持部に保持されている文書の論理構造のうちから前記構造パターン保持部に保持されている構造パターンに適合している部分を探索する処理を行なう構造パターンマッチ処理部を有し、該構造パターンマッチ処理部は、前記テンプレート保持部に保持されるテンプレートを利用して探索処理の範囲を論理構造の一部に限定し、限定した範囲内で探索処理を行なうことを特徴とする文書処理装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、構造化文書の検索、自動生成を行なう文書処理装置に関するものである。

【0002】

【従来の技術】 構造化文書を扱う文書処理装置において、文書の効率的な作成を支援するため、複数の文書を合成するなど、プログラムによって、新たな構造化文書を自動生成する文書自動生成処理や、図などの構造化文書の部品を検索し、部品単位での再利用を行なう文書部品検索処理などが行なえる。このような文書自動生成処理や文書部品検索処理では、構造パターンマッチが有効である。ここで、構造パターンマッチとは、指定された構造パターンに適合する文書部品を、文書構造の中から探すことをいう。文書自動生成処理では、文書から切り出す部品の位置、文書中に部品を挿入する位置を構造パターンによって指定する。文書部品検索では、文書部品の検索を構造パターンによって指示する。

【0003】 構造パターンとは、「文書論理構造中の部品に対する条件」と、「部品の接続関係にある条件」を組合わせたものである。例えば、「図表見出しに文字列『インストール』が表れる図」は、構造パターンの1例である。文書部品Aがこの構造パターンに適合するのは、

1. 文書部品Aは図である。
2. ある文書部品Bがあつて、
 - 2-1. 文書部品Bは文字列「インストール」を含む図表見出しである。
 - 2-2. 文書部品Bは文書構造中で、文書部品Aの直下にある。

という条件が成り立つときである。この例では、「文書論理構造中の部品に対する条件」が1. と2-1. の条件であり、「部品の接続関係にある条件」が2-2. の条件である。

【0004】 構造パターンマッチを用いている従来の文書処理システムとしては、例えば、DRAFT INTERNATIONAL STANDARD ISO/1

ECDIS 10179, "Information Technology-Text and office systems-Document Style Semantics and Specification Language (DSSSL)"がある。このDSSSLでは、構造パターンマッチのことをロケーションモデルと呼んでいる。

【0005】 また、情報処理学会第42回（平成3年前期）全国大会、6Q-4、「文書処理統合環境Dietにおける文書データ処理言語」に記載されているDPL (Diet Processing Language) においても、文書構造のパターンマッチを取り入れている。

【0006】 これらの文献においては、構造パターンマッチをいかにして行なうかは示されていない。例えば一つの方法として、文書構造を深さ優先に構造パターンの同じものを探索することにより、文書構造のパターンマッチが行なえることは容易に考えつく。しかし、この方法では、検索に時間がかかるという問題がある。

【0007】

【発明が解決しようとする課題】 本発明は、上述した事情に鑑みてなされたもので、効率よく構造パターンマッチを行なうことのできる文書処理装置を提供することを目的とするものである。

【0008】

【課題を解決するための手段】 本発明は、構造化文書の検索、自動生成を行なう文書処理装置において、文書の論理構造のテンプレートを保持するテンプレート保持部と、文書の論理構造を保持する論理構造保持部と、構造パターンを保持する構造パターン保持部と、前記論理構造保持部に保持されている文書の論理構造のうちから前記構造パターン保持部に保持されている構造パターンに適合している部分を探索する処理を行なう構造パターンマッチ処理部を有し、該構造パターンマッチ処理部は、前記テンプレート保持部に保持されるテンプレートを利用して探索処理の範囲を論理構造の一部に限定し、限定した範囲内で探索処理を行なうことを特徴とするものである。

【0009】

【作用】 論理構造保持部に保持される文書の論理構造は、テンプレート保持部に保持されるテンプレートに従っている。構造パターンマッチ処理部は、構造パターン保持部に保持される構造パターンと、テンプレート保持部に保持されているテンプレートを比較することによって探索の範囲を一部に限定することができる。この限定された探索の範囲内で、論理構造保持部に保持される文書の論理構造に対して構造パターンマッチを行なう。これにより、高速に構造パターンマッチを行なうことができる。

【0010】

【実施例】図1は、本発明の文書処理装置の一実施例を示すブロック図である。図中、1は論理構造保持部、2はテンプレート保持部、3は構造パターン保持部、4は構造パターンマッチ処理部である。論理構造保持部1は、文書の論理構造を保持する。テンプレート保持部2は、文書のテンプレートを保持する。論理構造保持部1に保持されている文書の論理構造は、テンプレート保持部2に保持されているテンプレートに基づいた構造を有している。構造パターン保持部3は、構造パターンマッチを行なう構造パターンを保持している。構造パターンマッチ処理部4は、論理構造保持部1に保持されている文書の論理構造のうちから、構造パターン保持部3に保持されている構造パターンに適合している部分を探索する。このとき、テンプレート保持部2に保持されている文書のテンプレートを用いて、文書の論理構造のうちから、探索処理を行なう範囲を限定し、限定された範囲内で探索処理を行なう。

論理根 ::= 梗概 章+ 文献リスト 付録 … (A)

梗概 ::= 段落 … (B)

章 ::= 章見出し 段落+ … (C)

文献リスト ::= 文献名+ … (D)

付録 ::= 図+ … (E)

ここで、“::=”は、右辺の項目の並びによって左辺が定義されていることを示し、右辺の項目の後ろの“+”は、その項目の1個以上の繰り返しを示している。

【0014】論理根は、文書を示す。文書は、式(A)のように定義され、1個の梗概、章の1個以上の繰り返し、1個の文献リスト、1個の付録から構成されている。梗概は、式(B)のように定義され、1個の段落から構成されている。章は、式(C)のように定義され、1個の章見出しと、段落の1個以上の繰り返しから構成されている。文献リストは、式(D)のように定義され、文献名の1個以上の繰り返しから構成されている。付録は、式(E)のように定義され、図の1個以上の繰り返しから構成されている。この文書の構造を図3に示している。

【0015】ここで、構造パターンとして、「図を持つ章」を考える。この構造パターンは、次の条件の組合わせである。

1. 文書部品Aは章である。
2. ある文書部品Bがあって、
 - 2-1. 文書部品Bは図である。
 - 2-2. 文書部品Bは文書構造中で文書部品Aに（直接または間接に）従属する。

【0016】構造パターンマッチ処理部4は、図2のステップ1において、文書テンプレートと構造パターンから次のように範囲の限定を行なう。まず、上述の式(A)乃至式(E)から、文書に出現し得る部品の種類は、論理根、章、文献リスト、付録、段落、章見出し、

【0011】図2は、構造パターンマッチ処理部における探索処理の流れを説明するためのフローチャートである。まず、ステップ1において、テンプレート保持部2に保持されているテンプレートと、構造パターン保持部3に保持されている構造パターンから、文書の論理構造のうちから、構造パターンマッチを行なう範囲を限定する。

【0012】次に、ステップ2において、論理構造保持部1に保持されている文書の論理構造全体のうち、ステップ1で決定した限定された範囲についてのみ、構造パターンマッチを行なう。これにより、部分的に構造パターンマッチを行なえばよいので、検索処理を高速化することができる。

【0013】構造パターンマッチ処理の具体例について述べる。図3は、文書テンプレートの一例の説明図である。文書テンプレートの一例として、次のようなものを考える。

文献名、図である。上述の条件1、「文書部品Aは章である。」を満たすのは、種類が章である部品のみである。また、上述の条件2-1、「文書部品Bは図である。」を満たすのは、種類が図である部品のみである。さらに、上述の条件2-2、「文書部品Bは文書構造中で文書部品Aに（直接または間接に）従属する」を満たすのは、次の組合わせに限られる。

A	B
論理根	梗概
論理根	章
論理根	文献リスト
論理根	付録
梗概	段落
論理根	段落
章	章見出し
章	段落
論理根	章見出し
論理根	段落
文献リスト	文献名
論理根	文献名
付録	図
論理根	図

【0017】この組み合わせを求める方法としては、例えば、推移的閉包の計算アルゴリズムとして広く知られているものを用いることができる。推移的閉包の計算アルゴリズムについては、例えば、Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, “Data Structures and Algorithms”: Addison Wesley等に記載されている。

【0018】上述のように、文書部品Aの種類が章で、文書部品Aに（直接または間接に）従属する文書部品B

5

であって、文書部品Bの種類が図であるものは存在しない。すなわち、上述の条件1.、2-1.、2-2を同時に満たす文書部品Aと文書部品Bの組み合わせはないことがわかる。従って、この例では、実際の文書の論理構造を調べることなく、構造パターンに適合する部分構造はないことがわかる。すなわち、構造パターンマッチ処理部4は、図2のステップ2でパターンマッチを行なう必要はない。

【0019】文書テンプレート中の式(C)が次のもので置き換えられたとしよう。

章：：＝章見出し (段落 | 図) +・・・(C')

ここで、(A|B)は、AまたはBの選択を示している。この式(C')は、章は1個の章見出しと、段落または図の1個以上の繰り返しからなることを表している。図4は、文書テンプレートの別の例の説明図である。図4では、ある1つの章が、1つの章見出しと1つの段落と1つの図を有する例を示している。

【0020】構造パターンマッチ処理部4は、図2のステップ1において、文書テンプレートと構造パターンから次のように範囲の限定を行なう。この例においても、上述の式(A)、(B)、(C')、(D)、(E)から、文書に出現し得る部品の種類は、論理根、章、文献リスト、付録、段落、章見出し、文献名、図である。上述の条件1.「文書部品Aは章である。」を満たすのは、種類が章である部品のみである。また、上述の条件2-1.「文書部品Bは図である。」を満たすのは、種類が図である部品のみである。さらに、上述の条件2-2.「文書部品Bは文書構造中で文書部品Aに(直接または間接に)従属する」を満たすのは、次の組合わせに限られる。

文書部品A	文書部品B
論理根	梗概
論理根	章
論理根	文献リスト
論理根	付録
梗概	段落
論理根	段落
章	章見出し
章	段落
章	図
論理根	章見出し
論理根	段落
論理根	図
文献リスト	文献名
論理根	文献名
付録	図

6

論理根 図

【0021】上述のように、この例の場合には、文書部品Aの種類が章で、文書部品Aに(直接または間接に)従属する文書部品Bであって、文書部品Bの種類が図であることがあり得ることがわかる。また、この組み合わせの結果から、章は梗概や文献リスト、付録の下位にはないことがわかる。従って、この例では梗概や文献リスト、付録の中は調べる必要はなく、章の中の図を探せば良いことがわかる。すなわち、図2のステップ2での構造パターンマッチは、章の中に対してのみ行なえば良く、梗概や文献リスト、付録の中に対しては行なわない。

【0022】上述の2つの具体例のように、文書テンプレートを用いて構造パターンマッチを行なう範囲を限定してから、文書の論理構造を用いて構造パターンマッチを行なうので、マッチングを行なう範囲を限定し、あるいは、先の例のように、構造パターンマッチを行なわなくてよい場合も検出することができる。

【0023】これらの具体例は、簡単な構造パターンおよび文書テンプレートについて示したが、もっと複雑な構造パターンや文書テンプレートであっても適用可能である。例えば、構造パターンとして、「図を持つ章であって、章題に『序文』という文字を有するもの」といった、より複雑な構造パターンについても処理可能である。

【0024】上述の文書テンプレートは、式により示したが、これは便宜上、式により表現しただけであり、他の表現方式を用いた文書テンプレートであっても、もちろんよい。

30 【0025】

【発明の効果】以上の説明から明らかなように、本発明によれば、文書テンプレートを用いて構造パターンマッチを行なう範囲を限定し、限定された範囲内で構造パターンマッチの処理を行なうので、構造パターンマッチの処理を効率良く、高速に行なうことができるという効果がある。

【図面の簡単な説明】

【図1】 本発明の文書処理装置の一実施例を示すブロック図である。

40 【図2】 構造パターンマッチ処理部における探索処理の流れを説明するためのフローチャートである。

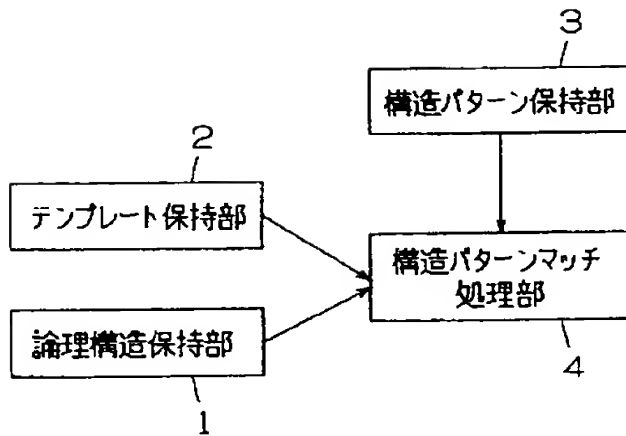
【図3】 文書テンプレートの一例の説明図である。

【図4】 文書テンプレートの別の例の説明図である。

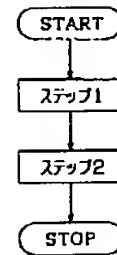
【符号の説明】

1 論理構造保持部、2 テンプレート保持部、3 構造パターン保持部、4 構造パターンマッチ処理部。

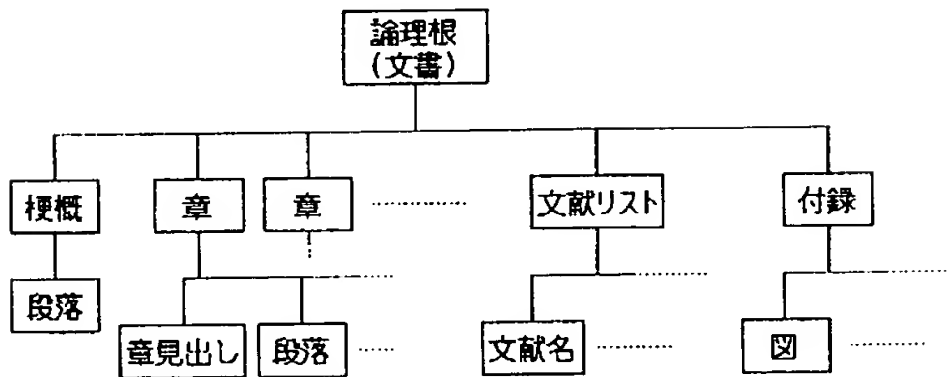
【図 1】



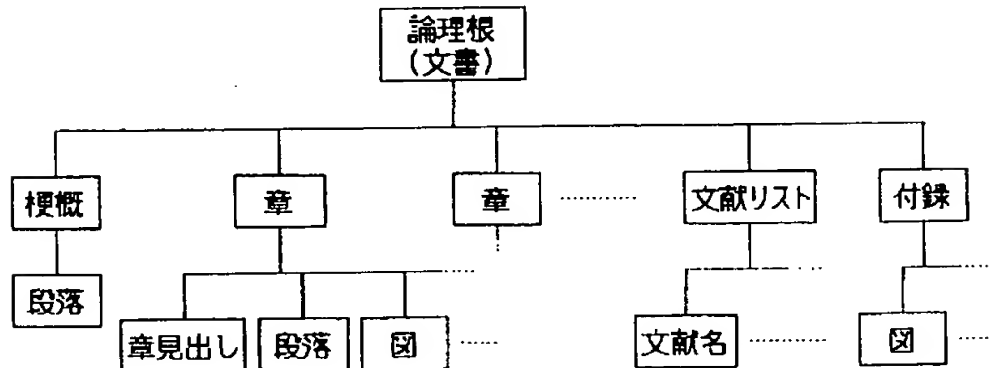
【図 2】



【図 3】



【図 4】



フロントページの続き